AP STATISTICS Ch. 12: Sample Surveys

***Vocabulary:***

**POPULATION-**  All experimental units that you want to make a conclusion about

**SAMPLING FRAME-** list of individuals from whom the sample is drawn. Not always the population of interest.

**SAMPLE –** small group of the population that you use in your experiment/study/survey.

**PARAMETER-**  describes a population. Often unknown. Fixed value.

**STATISTIC –** describes a sample of the population. Changes from sample to sample. We use the statistics from repeated samples to estimate the value of the parameter.

**VALUE PARAMETER STATISTIC**

Mean

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Std. Dev.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Proportion

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A sample is said to be **representative** **(or unbiased)** if the statistics accurately reflect the population parameter.

*EXAMPLE 1:*

*A polling agency takes a sample of 1500 American citizens from a list of tax returns and asks them if they are lactose intolerant.* ***12%*** *say yes. This is interesting, since it has been shown that* ***15%*** *of the population is lactose intolerant.*

*12% = \_\_\_\_\_\_\_\_\_ 15% = \_\_\_\_\_\_\_\_\_\_*

*Population? Sampling frame? Sample?*

*EXAMPLE 2:*

*A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted a year after they signed the card. It turned out that 210 (21%) of the sampled individuals had not smoked over the past six months.*

*21% = \_\_\_\_\_\_\_\_\_ Population =*

*Sampling frame= Sample =*

*Parameter of interest =*

*EXAMPLE 3:*

*On Tuesday, the bottles of tomato ketchup filled in a plant were supposed to contain an average of* ***14*** *ounces of ketchup. Quality control inspectors sampled 50 bottles at random from the day’s production. These bottles contained an average of* ***13.8*** *ounces of ketchup.*

*14 = \_\_\_\_\_\_\_\_\_ 13.8 = \_\_\_\_\_\_\_\_\_\_*

*Population? Sample? Sampling frame?*

*EXAMPLE 4:*

*A researcher wants to find out which of two pain relievers works better. He takes 100 volunteers and randomly gives half of them medicine #1 and the other half medicine #2.* ***17%*** *of people taking medicine 1 report improvement in their pain and* ***20%*** *of people taking medicine #2 report improvement in their pain.*

*17% = \_\_\_\_\_\_\_\_\_ 20% = \_\_\_\_\_\_\_\_\_\_*

*Population? Sampling frame? Sample?*

**BIAS VS. VARIABILITY**

**BIAS –** consistent, repeated measurements that are not close to the population parameter. Basically **accuracy.**

**VARIABILITY -**  basically like **reliability**. Consistent measurements (doesn’t matter if they are accurate or not.

* To reduce bias…
* To reduce variability …

**SAMPLING VARIABILITY –**

* Different samples give different results
* Different size samples give us different results
* Bigger samples are better!

**SAMPLING DISTRIBUTION-**  If we take lots of samples of the same size and make a histogram

**BIAS VS. VARIABILITY EXAMPLES:**

**UNBIASED ESTIMATOR-**  When the center of a sampling distribution (histogram) is equal to the true parameter

**SAMPLING DESIGNS**

**GOOD SAMPLING DESIGNS**

1. **Simple Random Sample (SRS)-** Every experimental unit has the same chance of being picked for the sample and every possible sample has the same chance of being picked.

Example: Take and SRS of 5 from the following list. Start at line 31 in the table.

Smith Jones Holloway

DeNizzo David Adams

Schaefer Gray Capito

Meyers Gingrich Card

Dietrich Moreland Hall

Walsh Whitter Jordan

Example: Take and SRS of 4 from the following list of math teachers. Start at line 18.

McGlone McCuen Wilson

Szarko Bellavance Woodring

Stotler Kelly Wheeles

Timmins Arden McNelis

Gemgnani O’Brien Robinson

Lorenz Lake Bainbridge

1. **STRATIFIED RANDOM SAMPLE (not SRS)-**

* Divide population into groups with something in common (called STRATA)
  + Example: gender, age, etc.
* Take separate SRS in each strata and combine these to make the full sample
  + Can sometimes be a % of each strata

Example: We want to take an accurate sample of CB South students. There are 540 sophomores, 585 juniors, and 530 seniors. Take a stratified sample.

1. **SYSTEMATIC RANDOM SAMPLE-**

The first experimental unit is selected at random, and each additional experimental unit is selected at a predetermined interval.

Examples:

1. **CLUSTER SAMPLE –**

Population is broken down into groups. All members of one (or more) group are taken as the sample.

1. **MULTI-STAGE SAMPLE-**

* Used for large populations
* Example: sampling the population of the USA:

**BIASED SAMPLING METHODS:**

1. **VOLUNTARY RESPONSE SAMPLES-**

Chooses itself by responding to a general appeal. Ex: call-in, write-in, etc.

1. **CONVENIENCE SAMPLES-**

Selecting individuals that are easiest to reach/contact

**TYPES OF BIAS IN A SAMPLE:**

* **UNDERCOVERAGE-**
* **NONRESPONSE-**
* **RESPONSE BIAS -**